Supreme Court, U.S.
FILED
SEP 14 1987
JOSEPH F. SPANIOL, JR.
CLERK

No. 86-6139

IN THE
# Supreme Court of the United States
OCTOBER TERM, 1987

---

CLARA WATSON,
*Petitioner,*

v.

FORT WORTH BANK & TRUST,
*Respondent.*

---

**On Writ of Certiorari to the United States Court of Appeals for the Fifth Circuit**

---

## BRIEF FOR *AMICUS CURIAE* AMERICAN PSYCHOLOGICAL ASSOCIATION IN SUPPORT OF PETITIONER

---

DONALD N. BERSOFF
(Counsel of Record)
LAUREL PYKE MALSON
DONALD B. VERRILLI, JR.
ENNIS FRIEDMAN & BERSOFF
1200 - 17th Street, N.W., Suite 400
Washington, D.C. 20036
(202) 775-8100
*Attorneys for* Amicus Curiae
*American Psychological Association*

September 14, 1987

WILSON - EPES PRINTING CO., INC. - 789-0096 - WASHINGTON, D.C. 20001

TABLE OF CONTENTS


| | Page |
|---|---|
| TABLE OF AUTHORITIES | iii |
| INTEREST OF *AMICUS CURIAE* | 1 |
| INTRODUCTION AND SUMMARY OF ARGUMENT | 2 |
| ARGUMENT | 4 |
| I. BECAUSE SUBJECTIVE ASSESSMENT DEVICES CAN, AND SHOULD, BE SCIENTIFICALLY VALIDATED, THE USE OF SUBJECTIVE SELECTION CRITERIA AND PROCEDURES BY EMPLOYERS SHOULD NOT PRECLUDE REVIEW UNDER ANY TITLE VII THEORY | 4 |
| A. Professional Standards Concerning the Technical Adequacy of Selection Devices are Applicable to the Subjective Methods Used by Respondent | 4 |
| B. There are Generally Accepted Strategies for Establishing the Validity of Subjective Methods of Employee Selection | 9 |
| C. To Reduce Sources of Bias the Validity of Each of the Selection Devices Used by Respondent Must and Can Be Established by Generally Accepted And Accessible Validation Strategies | 14 |
| 1. The interview | 14 |
| 2. Rating scales and other performance appraisals | 17 |
| 3. Experience requirements | 21 |
| II. THE SUBJECTIVE SELECTION PROCEDURES USED BY RESPONDENT FAIL TO MEET GENERALLY ACCEPTED STANDARDS AND APPEAR TO HAVE BEEN APPLIED WITHOUT ANY EVIDENCE THAT THEY ARE VALID FOR THE INFERENCES DRAWN FROM THEM | 22 |

TABLE OF CONTENTS—Continued

Page

III. THE FAILURE TO REQUIRE THAT SUBJECTIVE SELECTION DEVICES HAVE DEMONSTRABLE VALIDITY WOULD UNDERMINE THE PURPOSES OF TITLE VII- 28

CONCLUSION ---------- 30

## TABLE OF AUTHORITIES

CASES: Page

*Albemarle Paper Co. v. Moody*, 422 U.S. 405 (1975) ---- 20
*Brito v. Zia*, 478 F.2d 1200 (10th Cir. 1973) ---- 9
*Debra P. v. Turlington*, 644 F.2d 397 (5th Cir. 1981) ---- 6
*Douglas v. Hampton*, 512 F.2d 976 (D.C. Cir. 1975) ---- 6
*Griggs v. Duke Power Co.*, 401 U.S. 424 (1971) ---- 28, 29
*Harless v. Duck*, 14 FEP Cases 1616 (N.D. Ohio 1977) ---- 6
*McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973) ---- 29
*Texas Dep't of Comm. Affairs v. Burdine*, 450 U.S. 248 (1981) ---- 29
*Washington v. Davis*, 426 U.S. 227 (1976) ---- 6, 12
*Watson v. Fort Worth Bank & Trust*, 798 F.2d 791 (5th Cir. 1986) ---- 24, 27

STATUTES & REGULATIONS:

42 U.S.C. § 2000e *et seq.* ---- 2, 3, 28
29 C.F.R. § 1607 *et seq.* ---- 3, 7, 8, 27
H.R. Rep. No. 914, 88 Cong., 2d Sess., *reprinted in* 1964 U.S. Code Cong. & Ad. News 2391 ---- 28

MISCELLANEOUS:

AERA, APA, NCME, STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING (1985) ---- *passim*
AERA, APA, NCME, *Standards for Educational and Psychological Tests* (1974) ---- 6
A. ANASTASI, PSYCHOLOGICAL TESTING (5th ed. 1982) ---- 10, 15, 16, 17, 18, 20, 21
APA, *Standards for Educational and Psychological Tests and Manuals* (1966) ---- 6
APA, *Technical Recommendations for Psychological Tests and Diagnostic Techniques* (1954) ---- 6
Arvey, *Unfair Discrimination in the Employment Interview: Legal and Psychological Aspects*, 86 PSYCHOLOGY BULL. 736 (1982) ---- 14, 15

TABLE OF AUTHORITIES—Continued

Page

Arvey & Campion, *The Employment Interview: A Summary and Review of Recent Research*, 35 PERSONNEL PSYCHOLOGY 281 (1982) ---------- 16
Bernardin & Pence, *Effects of Rater Training*, 65 J. APPLIED PSYCHOLOGY 60 (1980) ---------- 21
Bersoff, *Testing and the Law*, 36 AM. PSYCHOLOGIST 1047 (1981) ---------- 11
W. BINGHAM, B. MOORE & J. GUSTAD, HOW TO INTERVIEW (4th ed. 1959) ---------- 15
Borman, *Format and Training Effects on Rating Accuracy and Rating Errors*, 64 J. APPLIED PSYCHOLOGY 410 (1979) ---------- 21
Brush & Owens, *Implementation and Evaluation for an Assessment Classification Model for Manpower Utilization*, 32 PERSONNEL PSYCHOLOGY 369 (1979) ---------- 22
Cascio & Bernardin, *Implications of Performance Appraisal Litigation for Personnel Decisions*, 34 PERSONNEL PSYCHOLOGY 211 (1981) ---------- 19, 20, 25
L. CRONBACH, ESSENTIALS OF PSYCHOLOGICAL TESTING (4th ed. 1984) ---------- 10, 18, 20
Distefano, Pryer, & Erffmeyer, *Application of Content Validity Methods to the Development of a Job-Related Performance Rating Criterion*, 36 PERSONNEL PSYCHOLOGY 621 (1983) ---------- 19
G. DREHER & P. SACKETT, PERSPECTIVES ON STAFFING AND SELECTION (1983) ---------- 21
Dunnette & Borman, *Personnel Selection and Classification Systems*, 40 ANN. REV. PSYCHOLOGY 477 (1979) ---------- 15
R. FEAR, THE EVALUATION INTERVIEW (2d ed. 1973) ---------- 15
Feild & Holley, *The Relationship of Performance Appraisal System Characteristics to Verdicts in Selected Employment Discrimination Cases*, 25 ACAD. MGMT J. 392 (1982) ---------- 19
Friedman & Williams, *Current Use of Tests for Employment in* 2 ABILITY TESTING (A. Wigdor & W. Garner eds. 1982) ---------- 9

TABLE OF AUTHORITIES—Continued

Page

S. GAEL, JOB ANALYSIS: A GUIDE TO ASSESSING WORK ACTIVITIES (1983) ---------- 13
I. GOLDSTEIN, TRAINING IN ORGANIZATIONS (2d ed. 1986) ----------5, 14, 18, 21
Grant & Bray, *Contributions of the Interview to Assessment of Management Personnel*, 53 J. APPLIED PSYCOLOGY 24 (1969) ---------- 15
Guion, *On Trinitarian Doctrines of Validity*, 11 PROF. PSYCHOLOGY 385 (1980) ---------- 11
Hakel, *Employment Interviewing in* PERSONNEL MANAGEMENT (K. Rowland & G. Ferris eds. 1982) ---------- 16
H. HENNEMAN, D. SCHWAB, J. FOSSUM, & L. DYER, PERSONNEL/HUMAN RESOURCE MANAGEMENT (1980) ---------- 14
Ivancevich, *Longitudinal Study of the Effects of Rater Training on Psychometric Errors in Ratings*, 64 J. APPLIED PSYCHOLOGY 502 (1979)---- 21
Kleiman & Durham, *Performance Appraisal, Promotion and the Courts: A Critical Review*, 34 PERSONNEL PSYCHOLOGY 103 (1981) ------5, 13, 19, 20
Korman, *The Prediction of Managerial Performance: A Review*, 21 PERSONNEL PSYCHOLOGY 295 (1968) ---------- 22
Kraiger & Ford, *A Meta-analysis of Ratee Race Effects in Performance Ratings*, 69 J. APPLIED PSYCHOLOGY 56 (1985) ---------- 18
Landy & Farr, *Performance Rating*, 87 PSYCHOLOGICAL BULL. 72 (1980) ----------17, 18, 19
A. LARSON & L. LARSON, EMPLOYMENT DISCRIMINATION § 15-87 (1986) ---------- 5
Latham, Saari, Pursell & Campion, *The Situational Interview*, 65 J. APPLIED PSYCHOLOGY 422 (1980) ---------- 21
Latham, Wexley & Pursell, *Training Managers to Minimize Rating Errors in the Observation of Behavior*, 60 J. APPLIED PSYCHOLOGY 550 (1975) ---------- 16
E. LEVINE, EVERYTHING YOU EVER WANTED TO KNOW ABOUT JOB ANALYSIS (1983) ---------- 13

TABLE OF AUTHORITIES—Continued

Page

Locher & Teel, *Performance Appraisal—A Survey of Current Practices*, 56 PERSONNEL J. 245 (1977) ..... 17

J. MATARAZZO & A. WEINS, THE INTERVIEW: RESEARCH ON ITS ANATOMY AND STRUCTURES (1972) ..... 15

E. MCCORMICK, JOB ANALYSIS: METHODS AND APPLICATIONS (1979) ..... 13

Messick, *Test Validity and the Ethics of Assessment*, 35 AM. PSYCHOLOGIST 1012 (1980) ..... 11

Owens, *Background Data in* HANDBOOK OF INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY (M. Dunnette ed. 1976) ..... 21

Owens & Schoenfeldt, *Toward a Classification of Persons*, 46 J. APPLIED PSYCHOLOGY 329 (1979) ..... 22

Pace & Schoenfeldt, *Legal Concerns in the Use of Weighted Applications*, 30 PERSONNEL PSYCHOLOGY 159 (1977) ..... 22

Reilly & Chao, *Validity and Fairness of Some Alternative Employee Selection Procedures*, 35 PERSONNEL PSYCHOLOGY 1 (1982) ..... 22

Rice, *Spotlight on Employee Performance*, 9 US AIR 53 (August 1987) ..... 19

Schmidt & Johnson, *Effect of Race on Peer Ratings in an Industrial Setting*, 57 J. APPLIED PSYCHOLOGY 237 (1973) ..... 18

Schmitt, *Social and Situational Determinants of Interview Decisions: Implications for the Employment Interview*, 29 PERSONNEL PSYCHOLOGY 79 (1976) ..... 14, 15, 23

B. SCHNEIDER & N. SCHMITT, STAFFING ORGANIZATIONS (2d ed. 1986) ..... *passim*

Schoenfeldt, *Utilization of Manpower: Development and Evaluation of Assessment-Classification Model for Matching Individuals with Jobs*, 59 J. APPLIED PSYCHOLOGY 583 (1974) ..... 22

TABLE OF AUTHORITIES—Continued

Page

SOCIETY FOR INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY, PRINCIPLES FOR THE VALIDATION AND USE OF PERSONNEL SELECTION PROCEDURES (1987) ---------- *passim*

Tenopyr, *Content-Construct Confusion*, 30 PERSONNEL PSYCHOLOGY 47 (1977) ---------- 11

Thompson & Thompson, *Court Standards for Job Analysis in Test Validation*, 35 PERSONNEL PSYCHOLOGY 865 (1982) ---------- 13, 23

Ulrich & Trumbo, *The Selection Interview Since 1949*, 63 PSYCHOLOGY BULL. 100 (1965) ---------- 14, 15

Waintroob, *The Developing Law of Equal Employment Opportunity at the White Collar and Professional Level*, 21 WM. & MARY L. REV. 45 (1979) ---------- 5

## BRIEF FOR *AMICUS CURIAE* AMERICAN PSYCHOLOGICAL ASSOCIATION IN SUPPORT OF PETITIONER

---

### INTEREST OF AMICUS CURIAE

The American Psychological Association ("APA") is a nonprofit, scientific, and professional organization with more than 65,000 members. It has been the major association of psychologists since 1892, and includes the vast majority of psychologists holding doctoral degrees from accredited universities in this country. Among APA's major functions are the promotion of psychological research, the dissemination of information regarding human psychological behavior, the promulgation of standards governing scientific and professional practice, including assessment, and, as reflected in its Bylaws, the "advance[ment] of psychology as a science and profession." A substantial number of APA's members are concerned with the development and validation of assessment devices for personnel selection in the employment context, including the more than 2500 members who belong to the APA's Division of Industrial and Organizational Psychology and the 1500 members who belong to its Division of Evaluation and Measurement.

The APA has participated as *amicus* in many cases in this Court involving social science issues, including *Kentucky v. Stincer*, 107 S. Ct. 2658 (1987) (effects on child victims of sex abuse of testifying in the presence of their alleged abusers); *Colorado v. Connelly*, 107 S. Ct. 515 (1986) (behavioral effects of command hallucinations); and *Lockhart v. McCree*, 106 S. Ct. 1758 (1986) ("conviction-proneness" of "death-qualified" juries). APA contributes *amicus* briefs only where it has special knowledge to share with the Court. APA regards this as one of those cases. In this instance, APA wishes to inform this Court of the state of current scientific thought regarding validation of personnel assessment devices, including subjective selection criteria and procedures such as those used by respondent in this case.

Petitioner and respondent have consented to the filing of this *amicus* brief. Their letters of consent are on file with the Clerk of the Court.

## INTRODUCTION AND SUMMARY OF ARGUMENT

APA addresses in this brief an essential and inherent issue in this case. *Amicus* will leave it to the parties to argue whether disparate impact analysis under Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e *et. seq.*, is properly applied to review the legality of subjective assessment devices for the hiring and promotion of employees. However, insofar as a negative answer is grounded in the assumption that subjective assessment devices are not amenable to psychometric scrutiny in the same way that ability tests are, such an assumption is contrary to fundamental and generally accepted scientific principles of measurement. The most frequently articulated reason for limiting disparate impact analysis to objective criteria and procedures—that only objective criteria and procedures yield sufficient statistical data to permit scientific validation—is not supported by the relevant social science literature. Indeed, the view that only objective selection criteria and procedures can be clearly identified, applied equally to all applicants, and statistically evaluated has been discredited by the extensive work of industrial psychologists and other assessment specialists. Subjective selection devices *can* be scientifically validated for the assessment of individuals for hiring, promotion, or other selection decisions in the employment context. The choice of analyses under Title VII, therefore, should not turn on whether the challenged employment practices are based on objective or subjective evaluations of applicants.

The APA's STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING (1985) [hereinafter STANDARDS] provide a framework for the evaluation and validation of testing and other assessment devices, including such subjective devices as interviews, behavioral observations, and rating scales. The STANDARDS are consistent with the PRINCIPLES FOR THE VALIDATION AND USE OF PERSONNEL

SELECTION PROCEDURES (1987) [hereinafter PRINCIPLES] published by the Society for Industrial and Organizational Psychology.[1] Furthermore, the Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. §§ 1607.1, *et. seq.* [hereinafter Uniform Guidelines] were explicitly intended to be consistent with the STANDARDS. *Id.* at § 1607.5(C). Such technical standards were clearly contemplated by the drafters of Title VII when they referred to the use of "professionally developed" assessment devices by employers. 42 U.S.C. § 2000e-2(h). When examined in light of the STANDARDS, PRINCIPLES, and Uniform Guidelines, it is clear that the procedures respondent used in this case to evaluate petitioner for promotion were not shown to be scientifically valid, *i.e.*, appropriate, meaningful, or useful for the inferences drawn from them. *See* Part I(B). More importantly, however, the STANDARDS and PRINCIPLES provide ample guidelines for how those procedures, notwithstanding their subjective nature, could have been validated.

The recognition that subjective assessment procedures may be validated is critical to the effectuation of the underlying goals of Title VII. Subjective procedures, like those of a more objective nature, should be required to be validated for specific jobs in any case where employers use those procedures as a defense to a prima facie case of discrimination, whether the claim is analyzed under a disparate treatment or a disparate effect theory. Only under such a rule will employers be inhibited from making personnel decisions based on unlawful and irrelevant factors. Should the Court in this case implicitly approve the use of unvalidated selection procedures as a defense to any Title VII claim, employers will have greater incentive to resort more readily to subjective assessment de-

---

[1] The Society is an integral component of the *amicus* and is also known as Division 14 of the APA (Division of Industrial and Organizational Psychology). Until adopted by *amicus* as a whole, the PRINCIPLES are the formal policy only of Division 14. However, they are "intended to represent the consensus of professional knowledge and thought as it exists today. . . ." *Id.* at 3.

vices, which would facilitate covert and discriminatory decisionmaking and severely undermine the right of equal employment opportunities for those classes of persons otherwise protected by Title VII.

## ARGUMENT

### I. BECAUSE SUBJECTIVE ASSESSMENT DEVICES CAN, AND SHOULD, BE SCIENTIFICALLY VALIDATED, THE USE OF SUBJECTIVE SELECTION CRITERIA AND PROCEDURES BY EMPLOYERS SHOULD NOT PRECLUDE REVIEW UNDER ANY TITLE VII THEORY.

*Amicus* relies on petitioner and other supporting *amici* to establish the applicability of disparate impact analysis under Title VII to subjective selection criteria for hiring and promotions. *Amicus* wishes to share with the Court its unique knowledge of testing and other assessment devices, and the validation of such devices, so that the Court's decision regarding the proper standard of analysis will be informed by relevant social science data. However, should the Court ultimately determine that disparate treatment analysis is the appropriate standard of review for subjective assessment devices in the employment context, *amicus* believes that such a holding in no way obviates the principle that such devices can and should be validated for the particular job in question.

#### A. *Professional Standards Concerning the Technical Adequacy of Selection Devices are Applicable to the Subjective Methods Used by Respondent.*

At issue in this case are three selection devices by which petitioner was evaluated for promotion by respondent's agents—interviews, supervisor's ratings, and experience requirements. Respondent characterizes these as subjective assessment procedures, in contrast to such procedures as multiple choice standardized paper-and-pencil tests which are typically classified as objective measures. Although in the context of personnel assessment procedures the term "subjective" is not easily defined, the concept has been used to refer variously to

procedures in which "judgment or discretion [is exercised] on the part of the evaluator" [2] or which lack any "neutral" factors,[3] to assessment devices of a "non-mechanical, operator-dependent" nature,[4] or to appraisals not based on " 'hard data,' such as production records [or] attendance." [5] Most simply, "Measures that require the statement of opinion, beliefs, or judgments are considered subjective." [6] Regardless of the particular definition, they are consistent with the widely-held view that subjective devices used by employers for hiring or promotion purposes are inherently less scientific, less quantifiable, less reliable, and less facially neutral than their objective counterparts.[7] For this reason, it has been as-

---

[2] Waintroob, *The Developing Law of Equal Employment Opportunity at the White Collar and Professional Level*, 21 WM. & MARY L. REV. 45, 48 (1979).

[3] 3 A. LARSON & L. LARSON, EMPLOYMENT DISCRIMINATION § 15-87 (1986).

[4] Brief of United States as *Amicus Curiae* on Petition for Writ of Certiorari at 13 [hereinafter Brief of United States].

[5] Kleiman & Durham, *Performance Appraisal, Promotion and the Courts: A Critical Review*, 34 PERSONNEL PSYCHOLOGY 103, 114 (1981) [hereinafter Kleiman & Durham].

[6] I. GOLDSTEIN, TRAINING IN ORGANIZATIONS 136 (2d ed. 1986) [hereinafter Goldstein]. "For example, rating scales are subjective measures, while measures of absenteeism are more objective. (However, supervisors' ratings of the absenteeism level of employees could turn that measure into a subjective criterion)." *Id.*

[7] "[S]ubjective measures are affected by the difficulties that one individual has in rating another without bias." *Id.* at 136-137. These difficulties, however, are not due to any inherent characteristics of subjective measures but to the failure of employers to apply standard principles of test construction to subjective measures. It is not intrinsic to such devices to be unquantifiable. When properly developed, they are amenable to scoring and objective analysis. For example:

> rating scales have been the most commonly employed measures in applied settings. . . . [One] reason is that it is simple to throw together a rating scale with a few traits . . . and delude yourself into believing that you have a useful measure of performance. Professionals . . . know that the *steps in*

serted that subjective selection methods and criteria are not susceptible to scientific "validation" or any other psychometric scrutiny. See, *e.g.*, Brief of United States, *supra* note 4, at 15, and references therein.

This view is fundamentally at odds with the universal judgment of those most experienced and knowledgeable in techniques of measurement and evaluation generally, and in the appraisal of employee performance specifically. The most authoritative source for the standards to be applied to determine the technical adequacy of assessment devices, the appropriateness of specific applications of these devices, and the reasonableness of inferences based on the results of these devices are the STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING (1985), a joint publication of the *amicus* APA, the American Educational Research Association ("AERA"), and the National Council on Measurement in Education ("NCME").[8]

Consistent with the STANDARDS, the Division 14 PRINCIPLES FOR THE VALIDATION AND USE OF PERSONNEL SE-

---

*the process are very similar for objective and subjective measures* and that shortcuts do not work in either case.

*Id.* at 137 (emphasis added).

[8] The 1985 STANDARDS represent the most modern expression of professional and scientific thinking concerning technical advances in psychological assessment. Its predecessor documents are, APA, *Technical Recommendations for Psychological Tests and Diagnostic Techniques* (1954); APA, *Standards for Educational and Psychological Tests and Manuals* (1966); and AERA, APA, NCME *Standards for Educational and Psychological Tests* (1974).

The 1966 and 1974 forerunners of the current STANDARDS have been cited with approval by this Court and in a variety of lower federal cases. *See, e.g., Washington v. Davis*, 426 U.S. 227, 247 n.13 (1976); *Debra P. v. Turlington*, 644 F.2d 397, 405 n.10 (5th Cir. 1981); *Douglas v. Hampton*, 512 F.2d 976, 984-986 (D.C. Cir. 1975) (see especially *id.* at 984 n.59 where the court called the 1966 edition "the universally recognized professional authority"); *Harless v. Duck*, 14 FEP Cases 1616, 1624 n.5 (N.D. Ohio 1977) (stating that the "courts have almost unanimously agreed that the

LECTION PROCEDURES (1987) apply the more general guidelines of the STANDARDS to the specific problems of making decisions in the context of employee selection, placement, and promotion and provide, *inter alia*, "principles for the application and use of valid selection procedures, and information that may be helpful to personnel managers and others responsible for authorizing or implementing validation efforts." PRINCIPLES at 2.[9]

Of relevance as well are the Uniform Guidelines. "The provisions of these guidelines . . . are intended to be consistent with generally accepted professional standards for evaluating standardized tests and other selection procedures, such as those described in the Standards . . . prepared by a joint committee of the" APA, AERA, and NCME. 29 C.F.R. § 1607.5(C). They "incorporate a single set of principles which are designed to assist employers . . . to comply with requirements of Federal law prohibiting employment practices which discriminate on the grounds of race. . . ." *Id.* at § 1607.1(B).

The STANDARDS, PRINCIPLES, and Uniform Guidelines apply to a broad range of selection procedures, not merely to the traditional objective measures commonly denominated as tests. Although the full title of the STANDARDS uses the word "testing," the term is generic and refers to "standardized ability . . . instruments, diagnostic and evaluative devices, interest inventories, personality inventories, and projective instruments." STANDARDS at 3. The term also includes samples of observable behavior "relevant to . . . employment decisionmaking." *Id.* at 4. *Amicus*, AERA, and NCME unequivocally view the

---

[1974] APA guidelines provide persuasive standards for evaluating claims of job relatedness").

[9] The 1987 PRINCIPLES are a revision of the original 1980 version. "The purposes of the revision are to bring the PRINCIPLES up to date scientifically, to make them consistent with the STANDARDS, and to reduce possible ambiguities regarding good practice in the use of selection procedures in making employment decisions." PRINCIPLES at 1.

STANDARDS as useful and applicable "to the entire range of assessment techniques." *Id.*[10]

Similarly, the Division 14 PRINCIPLES are not applicable solely to standardized paper-and-pencil tests. They are explicitly intended to aid employers who make hiring and promotion decisions to choose, select, develop, evaluate, and use all personnel selection devices, including "performance tests, . . . personality [and] interest inventories, . . . biographical data forms or scored application blanks, interviews, . . . experience requirements, . . . appraisals of job performance, . . . [and] estimates of advancement potential." PRINCIPLES at 1.

Finally, the Uniform Guidelines "provide a framework for determining" not only "the proper use of tests" but "other selection procedures" as well. 29 C.F.R. § 1607.1 (B). They plainly "apply to tests and other selection procedures which are used as a basis for any employment decision[,] including "hiring and promotion." *Id.* at § 1607.2(B).[11]

In sum, then, it is the universal professional judgment that the assessment devices used by respondent in deciding not to promote the petitioner can properly be scrutinized under the applicable scientific and professional standards, principles, and guidelines concerning the evaluation of the psychometric soundness of such devices.[12] In the present case, those devices were not sub-

---

[10] Such "instruments . . . are called tests here to indicate that the standards also apply to these instruments." *Id.* at 4-5.

[11] *See also, id.*, at § 1607.15(A)(1) (referring to selection procedures "either standardized or not standardized").

[12] *See, e.g.*, B. SCHNEIDER & N. SCHMITT, STAFFING ORGANIZATIONS 14 (2d ed. 1986) [hereinafter Schneider & Schmitt]:

> Typically we think of a test as an examination of some kind responded to with paper and pencil. . . . In fact, industrial psychologists and the Uniform Guidelines on Employee Selection Procedures (1978) have defined the word "test" in much broader terms and the courts have adopted this definition. In brief, a test is defined as any form of collecting information

jected to any such scrutiny to determine if they were valid for the purpose for which they were used.

### B. *There are Generally Accepted Strategies for Establishing the Validity of Subjective Methods of Employee Selection.*

There are many elements involved in the construction, development, and evaluation of an assessment instrument. Industrial psychologists and others who create a test or other selection procedure must choose the domain to be assessed, construct the items to which test takers will respond or select the behaviors to be observed, develop scoring scales and norms so that results can be interpreted, prepare manuals, and most importantly, ensure that the instrument is psychometrically sound. *See* STANDARDS at 9-37. The psychometric soundness of an instrument depends primarily on its being reliable [13] and valid. Although both qualities are essential, there is no doubt that "validity is the most important consideration in test evaluation." STANDARDS at 9.[14]

---

> when that information is used as a basis for making an employment decision. So, interviews are tests, as are application blanks, . . . performance appraisals used as a basis for making promotions (which, obviously are selection decisions), and any other kind of information used for making employment decisions.

*Accord* Friedman & Williams, *Current Use of Tests for Employment in* 2 ABILITY TESTING 99-100 (A. Wigdor & W. Garner eds. 1982) published by The National Academy of Sciences (acknowledging that the definition of selection procedures has extended to the full range of assessment devices, including interviews and are to be scrutinized according to the same guidelines used to evaluate standardized tests). *See also Brito v. Zia*, 478 F.2d 1200 (10th Cir. 1973) (holding that subjective observations by employer of minority employees had to be supported by empirical evidence of validity).

[13] "*Reliability* refers to the degree to which test scores are free from errors in measurement." STANDARDS at 19. The more reliable a test score the more consistent, dependable, or repeatable it will be. *See* PRINCIPLES at 39.

[14] "Undoubtedly the most important question to be asked about any psychological test concerns its validity . . . ." A. ANASTASI,

*Validation* refers to the process by which psychologists ascertain the degree to which certain inferences from a particular assessment device are appropriate, meaningful, or useful. *See* STANDARDS at 9; PRINCIPLES at 4. A test is valid if the proposed interpretation of scores proves to be sound and relevant. *See Cronbach*, *supra* note 14, at 125. Validity "concerns *what* the test measures and how well it does so. It tells us what can be inferred from test scores." Anastasi, *supra* note 14, at 131. When a selection procedure or device is said to be "validated," psychologists understand that the predictions inferred from the scored result of the procedure or device have a high rate of accuracy. For validation to be meaningful, it must predict performance of a *particular* task or set of tasks or other job relevant behavior of particular concern to the employer. Thus, selection procedures or devices are validated for a particular job when they have been demonstrated scientifically to make reliable and meaningful distinctions between individuals on the basis of their ability to perform particular tasks with competence or to function successfully in a particular job.[15] "The simple psychometric fact that test validity must be ascertained for specific uses of the test has long been familiar . . . . An invalid test or one that includes elements not related to the job under consideration may unfairly exclude minority group members who could have performed the job satisfactorily." Anastasi, *supra* note 14, at 432.

Validity "is a unitary concept." STANDARDS at 9. However, within the unifying theme of determining the

---

PSYCHOLOGICAL TESTING 27 (5th ed. 1982) [hereinafter Anastasi]. "Obviously, no aspect of a test is more important than validity . . . ." L. CRONBACH, ESSENTIALS OF PSYCHOLOGICAL TESTING 125 (4th ed. 1984) [hereinafter Cronbach]. The texts by Anastasi and Cronbach are considered the most authoritative basic treatises on the topic of psychological measurement.

[15] "Only as abbreviation is it legitimate to speak of 'the validity of a test'; a test relevant to one decision may have no value for another. So users must ask, 'How valid is this test for the decision made?' or 'How valid are the interpretations I am making of the scores?'." Cronbach, *supra* note 14, at 125.

relevance or interpretability of scored results, "[t]he validity of any inference can be determined in a variety of ways." PRINCIPLES at 4. "[T]he various means of accumulating validity evidence have been grouped into categories called content-related, criterion-related, and construct-related evidence of validity." STANDARDS at 9.[16] In the context of personnel selection procedures, content-validation involves a determination that the assessment instrument accurately reflects a representative sample of important aspects of job performance or job-required knowledge. *See* PRINCIPLES at 19.[17] Criterion-related validation involves a determination that the assessment instrument is predictive of, or significantly correlated with, important elements of job performance or work behavior. *See* PRINCIPLES at 6.[18] Construct validation

---

[16] "[I]nsofar as the courts have interpreted the test standards and . . . the Uniform Guidelines . . . to mean that content, criterion, and construct validity are distinct forms of validation, those interpretations are oversimplified, if not erroneous." Bersoff, *Testing and the Law*, 36 AM. PSYCHOLOGIST 1047, 1051 (1981). These three approaches should be viewed as subsets within a unifying and common framework. *See* Guion, *On Trinitarian Doctrines of Validity*, 11 PROF. PSYCHOLOGY 385, 386 (1980); Messick, *Test Validity and the Ethics of Assessment*, 35 AM. PSYCHOLOGIST 1012 (1980); Tenopyr, *Content-Construct Confusion*, 30 PERSONNEL PSYCHOLOGY 47 (1977). The three approaches may be discussed "separately only to acknowledge traditional presentations and avoid an abrupt departure from tradition." PRINCIPLES at 4.

[17] "In general, content-related evidence demonstrates the degree to which the sample of items, tasks, or questions on a test are representative of some defined universe or domain of content." STANDARDS at 10.

[18] "Criterion-related evidence demonstrates that test scores are systematically related to one or more outcome criteria . . . ." STANDARDS at 11.

> Two designs for obtaining criteria-related evidence—predictive and concurrent—can be distinguished. A predictive study obtains information about the accuracy with which early test data can be used to estimate criterion scores that will be obtained in the future. A concurrent study serves the same purposes, but obtains prediction and criterion information simultaneously.

*Id.*

involves a determination that the assessment instrument accurately measures the degree to which individuals possess identifiable characteristics which have been determined to be important for successful job performance. *See* PRINCIPLES at 25.[19]

Each of the validation strategies requires the employer to engage in an essential prerequisite activity. Through a process known as job analysis, the employer must clearly identify the most important components of successful job performance. "Job analysis is essential to the development of a content-oriented procedure or to the justification of a construct important to job behavior." PRINCIPLES at 5. "In some situations, the major purpose of job analysis may be to provide information from which criterion measures may be developed." *Id.* at 6. Satisfying this essential prerequisite requires an analysis of the job in question, and a clear articulation of the knowledge, skills and abilities ("KSAs"), or other personal characteristics or behaviors the exhibition of which determine proficiency at that job.[20] These data can be secured through judgments of job incumbents, their supervisors, personnel specialists, the professional judgment of job experts, and through training manuals,

---

[19] "The evidence classed in the construct-related category focuses primarily on the test score as a measure of the psychological category of interest . . . . Such characteristics are referred to as constructs because they are theoretical constructions about the nature of human behavior." STANDARDS at 9.

This Court has summarized these approaches in *Washington v. Davis*, 426 U.S. at 247 n.13, as it understood them to be described in the 1966 version of the STANDARDS.

[20] Although a job analysis is crucial to all validation strategies it has been emphasized more clearly in the context of content-oriented validity. *See, e.g.*, "Content validation should be based on a thorough and explicit definition of the content domain of interest. For job selection, classification, and promotion, the characterization of the domain should be based on job analysis." Standard 10.4, STANDARDS at 60-61. But it is required for criterion-related validity, *see* Standard 10.1, STANDARDS at 60, and construct-related validity, *see* Standard 10.8, STANDARDS at 61, as well.

job descriptions, and other written information. *See* PRINCIPLES at 19-20; Schneider & Schmitt, *supra* note 12, at 47; Thompson & Thompson, *Court Standards for Job Analysis in Test Validation*, 35 PERSONNEL PSYCHOLOGY 865 (1982). In addition, the relative importance of these KSAs must be determined. Finally, a close link between the assessment device and the identified job content or behavioral characteristic (construct) must be established. Standard 10.5, STANDARDS at 61; *see also* Standard 10.8. *Id.*[8]

In sum, then, the use of particular selection procedures by an employer reflects his or her implicit assumption that some important aspect of behavior on the job can be predicted from an individual's scores or performance on the chosen selection procedure. The critical factor underlying this assumption is the accumulation of evidence or data to support an inference of the chosen procedure's job-relatedness. This can be accomplished only through the various strategies of validation.

Validation is no less applicable to subjective assessment devices than to objective ones. In both cases, accurate predictors of job performance are essential to assist employers in selecting or promoting individuals who will best serve their needs, as well as to provide a method of personnel selection that inhibits consideration of non-job-related factors such as an individual's race. Indeed, several commentators have noted the emphasis placed by many courts and employers on objectivity, at the expense of validity, *i.e.*, requiring the use of certain "neutral appearing" objective tests to measure job performance, even where the validity of those criteria is clearly questionable.[22] Because of the role that validation plays in

---

[21] Several books summarize job analysis procedures and discuss their relative utility in various situations. *See, e.g.*, S. GAEL, JOB ANALYSIS: A GUIDE TO ASSESSING WORK ACTIVITIES (1983); E. LEVINE, EVERYTHING YOU EVER WANTED TO KNOW ABOUT JOB ANALYSIS (1983); E MCCORMICK, JOB ANALYSIS: METHODS AND APPLICATIONS (1979).

[22] *See, e.g.*, Kleiman & Durham, *supra* note 5, at 117-118.

enhancing the quality of selection procedures and reducing the potential for discrimination, an employer should be required to provide evidence of the validity of those procedures whether the employer chooses to use ones labeled as subjective or objective.

### C. *To Reduce Sources of Bias the Validity of Each of the Selection Devices Used by Respondent Must and Can Be Established by Generally Accepted and Accessible Validation Strategies.*

Industrial psychologists routinely use the three strategies for validating assessment devices described in Part I(B) to validate both objective devices such as standardized ability tests and interest inventories, and purely subjective or multi-component devices such as interviews, performance appraisal ratings, constructed performance tasks, nonscored experience and biographical data intake sheets, and structured behavioral sample tests. *See generally, e.g.*, Goldstein, *supra* note 6; Schneider & Schmitt, *supra* note 12.

#### 1. *The interview.*

The employment interview, the technique most heavily relied upon by respondent in this case, is probably more widely used than any other selection tool. *See* H. HENNEMAN, D. SCHWAB, J. FOSSUM, & L. DYER, PERSONNEL/HUMAN RESOURCE MANAGEMENT (1980); Ulrich & Trumbo, *The Selection Interview Since 1949*, 64 PSYCHOLOGY BULL. 100 (1965). However, because most employers are unaware of research which has determined which variables are reliably, validly, and uniquely assessed in the selection interview, *see* Schmitt, *Social and Situational Determinants of Interview Decisions: Implications for the Employment Interview*, 29 PERSONNEL PSYCHOLOGY 79, 97 (1976), the employment interview is typically subject to interview bias of various types. *See* Arvey, *Unfair Discrimination in the Employment Interview: Legal and Psychological Aspects*, 86 PSYCHOLOGY BULL. 736 (1982). The most commonly known bias is "halo effect," where an interviewer may be unduly influ-

enced by a single trait which colors his/her judgment of the employee's other traits. *See* Anastasi, *supra* note 14, at 612. A related problem is "stereotyping" in which an employee is judged "based on his or her group membership [e.g., race] rather than on the basis of his or her unique characteristics" Schneider & Schmitt *supra* note 12, at 388-389. Another concern is the "similar-to-me phenomenon" in which the interviewer adopts the attitude "I am wonderful and I have the following attitudes and opinions, so if candidates I interview have the same attitudes and opinions, they must also be wonderful." *Id.* at 389. "When combined with stereotyping, the similar-to-me phenomenon can be a potent determinant of interviewer decision-making." *Id.*

Interviews can afford an opportunity for direct observation of samples of behavior, albeit limited, manifested during the interview and serve to evoke life-history data, both of which can be important predictors of future performance, if interviews are developed and conducted using generally accepted standards. *See* Anastasi, *supra* note 14, at 610.[23] Several recent studies, reviewed by noted scholars, show that interview judgments can be valid indicators of subsequent job performance. *See*

---

[23] A variety of available sources discuss methods, applications, and effectiveness of interviewing, and research on the interviewing process. *See, e.g.*, W. BINGHAM, B. MOORE & J. GUSTAD, HOW TO INTERVIEW (4th ed. 1959); R. FEAR, THE EVALUATION INTERVIEW (2d ed. 1973); J. MATARAZZO & A. WEINS, THE INTERVIEW: RESEARCH ON ITS ANATOMY AND STRUCTURE (1972); Arvey, *Unfair Discrimination in the Employment Interview: Legal and Psychological Aspects*, 86 PSYCHOLOGY BULL. 736 (1979); Dunnette & Borman, *Personnel Selection and Classification Systems*, 40 ANN. REV. PSYCHOLOGY 477 (1979); Grant & Bray, *Contributions of the Interview to Assessment of Management Personnel*, 53 J. APPLIED PSYCHOLOGY 24 (1969); Schmitt, *Social and Situational Determinants of Interview Decisions: Implications for the Employment Interview*, 29 PERSONNEL PSYCHOLOGY 79 (1976); Ulrich & Trumbo, *The Selection Interview Since 1949*, 63 PSYCHOLOGY BULL. 100 (1965).

Arvey & Campion, *The Employment Interview: A Summary and Review of Recent Research*, 35 PERSONNEL PSYCHOLOGY 281 (1982). These sources show that interviews can be created that are valid and nondiscriminatory if interview questions are carefully linked to job analysis and performance criterion data.[24]

Interview validity is not alone sufficient, however. The validity of the *interviewer* must also be established. An "interview requires skill in data gathering and in data interpreting. An interview may lead to wrong decisions because important data were not elicited or because given data were inadequately or incorrectly interpreted." Anastasi, *supra* note 14, at 610-611. In this regard, interviewer training is important. A structured interview guide will improve interviewer reliability and assist in removing any bias especially if the training occurs with applicants of gender and/or race different than that of the interviewer. With important interviews used to determine hiring or promotion, it is very helpful if applicants are seen by more than one interviewer, although if records are kept it is possible to identify those interviewers whose decisions are most reliable and valid and rely on those interviewers singly to make judgments. *See* Schneider & Schmitt, *supra* note 12, at 390-394; Hakel, *Employment Interviewing in* PERSONNEL MANAGEMENT (K. Rowland & G. Ferris eds. 1982).

---

[24] As one example, researchers have described and tested an innovative but relatively simple and valid employment interview. Critical incidents, *i.e.*, reports by job incumbents or supervisors of situations in which especially effective or ineffective behavior is displayed, were converted into situational interview questions. The interviewer posed these situations to job applicants and asked them how they would behave. Each answer was rated independently by two or more interviewers on a five-point scale with end points on the scale provided by job experts to facilitate objective scoring. The process validily predicted future job performance for both women and blacks. *See* Latham, Saari, Pursell & Campion, *The Situational Interview*, 65 J. APPLIED PSYCHOLOGY 422 (1980).

In sum, the use of employment interviews should be "preceded by a thorough analysis of the target job, the development of a structured set of questions based on the job analysis, and the development of behaviorally specific rating instruments by which to evaluate applicants." Schneider & Schmitt, *supra* note 12, at 395. The assessment of employees "should be maximally dependent on their personal characteristics and minimally dependent on who made the assessment . . . . Where non-test predictors like interviewer judgments are used, the [employer] should develop procedures that will minimize error resulting from differences between judges." PRINCIPLES at 12.

**2. *Rating scales and other performance appraisals.***

Performance appraisal devices such as rating scales are widely used in employment settings, and were used by respondent in this case.[25] "Rating scales differ from naturalistic observations in that data are accumulated casually and informally; they also involve interpretation and judgment, rather than simple recording of observations." Anastasi, *supra* note 14, at 611. In contrast to interviews, however, "they typically cover a longer observation period and the information is obtained under more realistic conditions." *Id.*

Like interviews, rating scales are subject to a variety of sources of contamination or bias, including: (1) *Opportunity bias* which occurs if raters do not have the opportunity to observe the employee in situations in which the behavior to be rated could be manifested, but have the opportunity to do so with a competing employee; (2) *halo effect*, a tendency on the part of raters to be unduly influenced by a single favorable or unfavorable

---

[25] In one survey, 89% of the companies studied reported using performance appraisals on a regular basis. Locher & Teel, *Performance Appraisal—A Survey of Current Practices*, 56 PERSONNEL J. 245 (1977). Of all performance appraisal techniques, the rating scale is by far the most ubiquitous. *See* Landy & Farr, *Performance Rating*, 87 PSYCHOLOGICAL BULL. 72, 73 (1980) [hereinafter *Performance Rating*].

trait, which colors their judgment on the individual's other traits; (3) *error of central tendency*, or the tendency to place persons in the middle of the scale and to avoid extremes; and (4) *leniency error*, or the reluctance of many raters to assign unfavorable ratings. Both latter errors reduce the effective width of the rating scale and make them less useful in distinguishing among individuals. *See Anastasi*, *supra* note 14, at 611-612; Cronbach, *supra* note 14, at 509-511; Schneider & Schmitt, *supra* note 12, at 90-92; *Goldstein*, *supra* note 6, at 255.

Most troubling in the context of this case is that scores on rating scales may be affected by the race of the rater and ratee. White raters have been found to assign significantly higher ratings to white ratees than black ratees. These findings were noted in a comprehensive review of 74 studies involving 17,159 ratees in which the rater was white and 14 studies involving 2,420 ratees in which the rater was black. Race effects were more pronounced in real-life settings than in laboratory settings and more likely when, as in this case, the proportion of blacks in the workforce was small. *See* Kraiger & Ford, *A Meta-analysis of Ratee Race Effects in Performance Ratings*, 69 J. APPLIED PSYCHOLOGY 56 (1985).[26]

Psychologists have published a great deal of accessible literature describing rating scale formats which are useful in rating employees and have been critiqued in the

---

[26] These findings are not universal nor do they imply that performance appraisal systems are inherently discriminatory. One study, for example, found no race of rater effect in an industrial setting which was highly racially integrated and where participants in the study had been exposed to human relations training. *See* Schmidt & Johnson, *Effect of Race on Peer Ratings in an Industrial Setting*, 57 J. APPLIED PSYCHOLOGY 237 (1973). These mitigating factors are not true with regard to respondent. For a comprehensive review of the effects of rater and ratee characteristics and the interaction of the two *see Performance Rating*, *supra* note 25, at 74-82.

context of Title VII requirements.[27] Many of these formats would be significant improvements over the system used by respondent.[28] Regardless of format, what does produce a superior scale is that it is the "result of psychometric rigor in development and of some level of participation of individuals representative of those who will eventually use the scales to make ratings . . . ." *Performance Rating*, *supra* note 25, at 85.

An essential aspect of the requirement for psychometric rigor is the job analysis. "The development of rating procedures should ordinarily be guided by job analyses

---

[27] *E.g.*, Cascio & Bernardin, *Implications of Performance Appraisal Litigation for Personnel Decisions*, 34 PERSONNEL PSYCHOLOGY 211 (1981); Distefano, Pryer, & Erffmeyer, *Application of Content Validity Methods to the Development of a Job-Related Performance Rating Criterion*, 36 PERSONNEL PSYCHOLOGY 621 (1983); Feild & Holley, *The Relationship of Performance Appraisal System Characteristics to Verdicts in Select Employment Discrimination Cases*, 25 ACAD. MGMT J. 392 (1982); Kleiman & Durham, *supra* note 5; *Performance Rating*, *supra* note 25. The dissemination of this information is so widespread that it now appears in popular literature designed for lay readers. *See* Rice, *Spotlight on Employee Performance*, 9 US AIR 53 (August 1987).

[28] Two of the most popular rating formats are the graphic rating scale and the behaviorally anchored rating scale ("BARS"). In the graphic rating scale's most usual format, several dimensions to be rated are listed vertically and raters are then asked to make rating decisions along a horizontal 5 to 9 point scale. For example, if the dimension to be rated is "accuracy," the scale may use a numerical or one-word verbal rating, *e.g.*, 1 to 5 or "high" to "low"; preferably, it may use a range of descriptions, *e.g.*, at one end of the scale, would be "makes too many errors"; at the other end, "almost never makes mistakes." BARS uses dimensions derived by raters who would actually use the scale with different points on each dimension anchored by statements describing actual job behavior which would illustrate specific levels of performance, *e.g.*, using "accuracy" in rating a bank teller, the statements could include a range from "makes frequent errors in totalling accounts at end of day" to "errors in totalling accounts are consistently rare." *See generally* Schneider & Schmitt, *supra* note 12, at 101-106. Neither scale format seems more useful than the other in practice. *See Performance Rating*, *supra* note 25, at 85.

if, for example, raters are expected to evaluate several different aspects of performance," as in this case. PRINCIPLES at 10. Job analyses are also important if, as here, appraisal of past performance is used as a predictor of future performance. The use of ratings of past performance in one job to make promotion decisions for another position is only permissible if the ratings of past performance are valid and the ratings of past performance are related to future performance. The latter requires a job analysis indicating the extent to which the two jobs overlap. *See* Cascio & Bernardin, *Implications of Performance Appraisal Litigation for Personnel Decisions*, 34 PERSONNEL PSYCHOLOGY 211, 217 (1981).[29]

The usefulness of a rating scale is also highly dependent on the skill of the rater. "Valid ratings cannot be made by someone who is either unfamiliar with the work of the ratee or lacks the skills necessary to accurately observe or rate the job behavior." Kleiman & Durham, *supra* note 5, at 113. In these respects, "those in immediate contact with the subject give superior information," Cronbach, *supra* note 14, at 512 (in the case of employment settings, first level supervisors), and those raters who have undergone training show increased "reliability and validity of ratings" and decreased errors in judgment. Anastasi, *supra* note 14, at 612.[30]

---

[29] *See also Albemarle Paper Co. v. Moody*, 422 U.S. 405, 431-433 (1975) where this Court, invoking the Uniform Guidelines (and crediting APA's STANDARDS) condemned as "materially defective" the employer's validation study because its "subjective supervisorial rankings" used standards which were vague and ambiguous and failed to follow the Uniform Guidelines' requirement for job analyses.

[30] *See also* Standard 1.13, STANDARDS at 16: "When criteria are composed of rater judgments, the degree of knowledge that raters have concerning ratee performance should be reported. If possible, the training and experience of the raters should be described"; PRINCIPLES at 10: "It may . . . be necessary to train raters in the observation and evaluation of performance. Further, supervisors should be expected to be familiar enough with the demands of the job to evaluate overall performance." The utility of rater training

In sum, rating scales will conform to legal and psychometric requirements if the appraisal system is based on a job analysis, contains clearly defined dimensions of job performance rather than vague, global measures or abstract trait names, is behaviorally based so that all ratings can be supported by objective, observable evidence, and if the raters are in the position to observe the behaviors to be rated and are trained to reduce sources of bias, contamination, or other rating errors.

3. *Experience requirements.*

The use of past experience to make judgments about future performance, as in this case, is one aspect of a recognized selection device called biographical inventory technique or, more commonly, "biodata." *See* Owens, *Background Data in* HANDBOOK OF INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY (M. Dunnette ed. 1976). When used properly, biographical inventories which include prior experience are "especially appropriate for assessing the qualifications of women and minority groups." Anastasi, *supra* note 14, at 616. To serve this purpose, however, the biodata inventory must focus on "specific, job-relevant past achievements, rather than on the passive exposure implied by the customary education and experience records." *Id.* Respondent's use of an experience criterion falls far short of this standard.

In fact, use of biodata is relatively likely to produce adverse impact if biodata items are not chosen carefully. *See* G. DREHER & P. SACKETT, PERSPECTIVES ON STAFFING AND SELECTION (1983). A number of comprehensive

---

in reducing rating errors and minimizing bias, as well as providing employers with useful techniques for doing so, is demonstrated in, *e.g.*, Bernardin & Pence, *Effects of Rater Training*, 65 J. APPLIED PSYCHOLOGY 60 (1980); Borman, *Format and Training Effects on Rating Accuracy and Rating Errors*, 64 J. APPLIED PSYCHOLOGY 410 (1979); Ivancevich, *Longitudinal Study of the Effects of Rater Training on Psychometric Errors in Ratings*, 64 J. APPLIED PSYCHOLOGY 502 (1979); Latham, Wexley & Pursell, *Training Managers to Minimize Rating Errors in the Observation of Behavior*, 60 J. APPLIED PSYCHOLOGY 550 (1975). *See generally* Goldstein, *supra* note 6, at 254-259.

validity studies have been conducted on the use of biodata as a selection device, concluding that it has inconsistent validity even in a form more comprehensive than used by respondent. *See, e.g.*, Reilly & Chao, *Validity and Fairness of Some Alternative Employee Selection Procedures*, 35 PERSONNEL PSYCHOLOGY 1 (1982).[31]

However, as with all selection devices, the most reasonable and justifiable approach in using biodata is to base the choice of items on a well done job analysis, matching the items to the knowledge, skill, and ability requirements of the job description. *See* Pace & Schoenfeldt, *Legal Concerns in the Use of Weighted Applications*, 30 PERSONNEL PSYCHOLOGY, 159 (1977). Past experience alone, without the careful selection of both logically and empirically justified life-history questions, is unacceptable as a selection device.[32]

## II. THE SUBJECTIVE SELECTION PROCEDURES USED BY RESPONDENT FAIL TO MEET GENERALLY ACCEPTED STANDARDS AND APPEAR TO HAVE BEEN APPLIED WITHOUT ANY EVIDENCE THAT THEY ARE VALID FOR THE INFERENCES DRAWN FROM THEM.

When reviewed in the context of the principles and studies described in Part I, it is clear that the assessment

---

[31] *Compare* Reilly & Chao, *Validity and Fairness of Some Alternative Employee Selection Procedures*, 35 PERSONNEL PSYCHOLOGY 1 (1982) (satisfactory reliability) with Korman, *The Prediction of Managerial Performance: A Review*, 21 PERSONNEL PSYCHOLOGY 295 (1968) (finding biodata to have lower validity than other predictors for predicting managerial performance).

[32] Sophisticated research on the use of valid biodata is available to inform the interested employer. *See, e.g.*, Brush & Owens, *Implementation and Evaluation for an Assessment Classification Model for Manpower Utilization*, 32 PERSONNEL PSYCHOLOGY 369 (1979); Schoenfeldt, *Utilization of Manpower: Development and Evaluation of Assessment-Classification Model for Matching Individuals with Jobs*, 59 J. APPLIED PSYCHOLOGY 583 (1974); Owens & Schoenfeldt, *Toward a Classification of Persons*, 46 J. APPLIED PSYCHOLOGY 329 (1979). *See generally* Schneider & Schmitt, *supra* note 12, at 378-382.

devices used by respondent in this case to select employees, including petitioner, for promotion to the position of teller supervisor are distressingly inadequate. They have been used to deny promotion to a member of a protected class without any evidence that they were developed, used, and applied in a way consistent with generally accepted professional standards. There is absolutely no evidence that the procedures were subjected to any of the validation strategies available to respondent, even in rudimentary form.

It is not unlawful *per se* for employers to use so-called "subjective" selection procedures. *See* Part I(A). But, assuming that the use of subjective criteria was appropriate for the position of supervisor of tellers, respondent failed to perform a job analysis for the position to identify more accurately the knowledges, skills, and abilities which are desirable for successful performance of the job.[33] There is no evidence that respondent secured accurate and thorough information about the job from job incumbents, their supervisors, personnel specialists, training manuals, job descriptions, or actual observation by trained observers. *See* PRINCIPLES at 5-6, 19-24; Schneider & Schmitt, *supra* note 12, at 47-50.

With regard to the selection procedures themselves, there are crucial infirmities in each of them. With regard to the *interview*, there is no evidence that the interviewer, in this case a white male, had any training in conducting interviews, a requirement that is especially important when the interviewee is of a different race and gender than the interviewer.[34] Nor was more than one

---

[33] As noted in Part I, job analysis is a critical first step in establishing the usefulness of any selection procedure. For a review of 26 employment discrimination cases yielding a helpful summary of requirements for judicially-approved job analyses, *see* Thompson & Thompson, *Court Standards for Job Analysis in Test Validation*, 35 PERSONNEL PSYCHOLOGY 865 (1982).

[34] "The training of interviewers especially with possible applicants of different race or sex may increase 'their ability to relate' . . . ." Schneider & Schmitt, *supra* note 12, at 386; *see* Schmitt, *Social and Situational Determinants of Interview Decisions: Impli-*

interviewer systematically involved in decisionmaking. There is also no evidence that the interviews were structured so as to improve reliability and reduce biasing errors,[35] nor is there any evidence that the nature of the interview or the questions asked had any empirical, logical, or theoretical connection with the position for which petitioner was considered.[36]

With regard to the *rating scales*, it was especially important for respondent to have offered some evidence for the validity of its supervisor performance appraisal, as the use of numerical values gives it facial validity and the appearance of objectivity. But, when judged according to the criteria discussed in Part I(B) & (C)(2) the rating scale is deplorable.

The qualities measured on the rating scale are neither unambiguously defined nor is there any demonstrable correlation between many of the criteria listed in the scale and successful job performance.[37] For example, all

---

*cations for the Employment Interview*, 29 PERSONNEL PSYCHOLOGY 79, 97 (1976). *See also* PRINCIPLES at 33: "All persons within the organizations who have responsibilities related to the use of employment tests and related predictors should be qualified through appropriate training to carry out their responsibilities."

[35] "Use of a structured interview guide will improve interviewer reliability." Schneider & Schmitt, *supra* note 12, at 386.

[36] "Predictor variables should be chosen for which there is an empirical, logical, or theoretical foundation." PRINCIPLES at 11.

[37] The qualities rated were accuracy of work, alertness, personal appearance, supervisor-coworker relations, quantity of work, physical fitness, attendance, dependability, stability ("the ability to withstand pressure and remain calm in most situations"), drive ("ambition"), friendliness and courtesy, and job knowledge. The qualities were variously rated on a scale from 0 to 7-10. *See Watson v. Fort Worth Bank & Trust*, 798 F.2d 791, 812 n.26 (5th Cir. 1986) (Goldberg, J., dissenting). "Few of these categories have much objective content. For example, 'personal appearance,' 'drive,' and 'friendliness and courtesy' are clearly subjective on their face. . . . The rating system is also subjective: [e.g.,] 0-1 'does not meet minimum requirement . . .; 7-8, 'superior work production record.' This type of subjective measurement lends itself to discriminatory bias, be it conscious or unconscious." *Id.*

but two of the rating criteria used by respondent are totally undefined and the two that have purported definitions (stability and drive) are only vaguely defined; none of qualities assessed have anchoring or endpoint definitions, as is customary in generally accepted graphic or BARS scales.[38] The scale thus failed to use clearly defined individual components or dimensions of job performance, in contrast to undefined global measures, *e.g.*, "neat and clean in appearance" vs. "personal appearance." [39] Similarly, it did not use behaviorally based performance dimensions that could be verified by objective, observable evidence, *e.g.*, "knows how to check account balance vs. "job knowledge." [40]

The failure to conduct even a rudimentary job analysis further undermines the rating scale's validity. It is far from clear how criteria such as "physical fitness," measure an individual's ability to supervise tellers. The constructs or behavior traits identified by respondent such as "drive" or "dependability" could be validated for use in promoting individuals to supervisory teller positions if demonstrated to be job-related and assessed reliably from the performance appraisal. Although the use of such "abstract trait names" is not advised unless the traits can be defined in terms of observable behaviors,[41] it may be necessary to measure such personality constructs for certain jobs. If certain traits or constructs are deemed important enough to influence personnel selection, however, they are important enough to measure validly:

> Knowing whether a construct is measured validly requires, if not a theory, at least some fairly well articulated ideas about what is being measured, what a

---

[38] *See supra* note 28.

[39] Cascio & Bernardin, *Implications of Performance Appraisal Litigation for Personnel Decisions*, 34 PERSONNEL PSYCHOLOGY 211, 212 (1981) [hereinafter Cascio & Bernardin].

[40] *Id.*

[41] Cascio & Bernardin, *supra* note 39, at 212.

> measure of the construct should reasonably be expected to be related to and, perhaps more importantly, what it should not be related to. . . . This view of constructs and construct validity implies two aspects of a construct-related strategy for developing evidence to judge the job relatedness of a selection procedure. The first is evidence that the construct is indeed important for job performance. . . . Ordinarily, a job analysis can provide a part of the basis for identifying and defining constructs which are important to job performance. Clarity of the articulation of the meaning and the nature of the construct, and well-informed expert judgment that a logical relationship exists between the nature of the construct and identifiable demands of the job is essential.
>
> The second is evidence that the instrument used as a selection procedure is a valid measure of the construct and not of other constructs.

PRINCIPLES at 25; *see also* STANDARDS 9-10.

Finally, there is no indication that the supervisors who performed the ratings had any training in assessing responses, avoiding the variety of sources of contamination of bias, especially when there is significant evidence that scores on rating scales may be affected by the race of rater and ratee. *See* Part I(C)(2).

With regard to *experience requirements*, respondent failed to show that the lack of prior experience upon which it based its failure to promote petitioner in her first three attempts was related to the position for which she applied. It could very well be that her experience as a teller would be predictive of her success as a supervisor of tellers, but without an analysis of both positions, that assumption could very well be faulty. In addition, respondent idiosyncratically employed the criterion, using her alleged lack of experience to justify its denying her promotion in three instances and then ignored the criterion to deny her promotion in a fourth instance, when it promoted a competing applicant with less experience. Perhaps most importantly, respondent failed to use prior

experience as one part of a carefully selected and logically and empirically justified comprehensive biographical data inventory.

Although respondent may have fewer resources to devote to assessment device development or validation than larger organizations, that fact does not excuse the absence of any attempt to support its use of the selection devices employed in this case:

> Where resources or sample sizes are limited, the criterion-related evidence of validity and content-related validation judgments obtained on similar jobs in other settings and the strength of the construct-related evidence of validity already generated by the [particular instrument] become particularly important. Employers should not be precluded from using a [particular instrument] if it can be demonstrated that th[at instrument] has generated a significant record of validity in similar job settings for highly similar people[,] or that it is otherwise appropriate to generalize from other applications.

STANDARDS at 59.[42] Respondent neglected to conduct even this inexpensive inquiry. There is absolutely no legitimate reason why respondent failed to conduct even crude validity studies of the selection devices it used to evaluate petitioner or, at the very least, to investigate the availability of existing sources of valid selection devices. These failures are all the more disturbing considering its poor record in hiring and promoting minority employees.[43]

---

[42] Researchers and employers are encouraged to conduct cooperative studies when adequate data . . . are not available." PRINCIPLES at: The Uniform Guidelines also clearly permit exceptions to the general requirement that employers validate their own procedures. *See* 29 C.F.R. §§ 1607.6-.8.

[43] "[P]laintiff presented 'significant proof' that the bank operated under a 'general policy of discrimination'." *Watson v. Fort Worth Bank & Trust*, 798 F.2d 791, 807 (5th Cir. 1986) (Goldberg, J., dissenting). *See, id.* at 810-814 for supporting statistical data.

## III. THE FAILURE TO REQUIRE THAT SUBJECTIVE SELECTION DEVICES HAVE DEMONSTRABLE VALIDITY WOULD UNDERMINE THE PURPOSES OF TITLE VII.

The underlying goal of Title VII of the Civil Rights Act of 1964 was the "eliminat[ion] . . . [of] discrimination in employment" based on race, color, religion, sex or national origin in all of its forms. H.R. Rep. No. 914, 88 Cong., 2d Sess., *reprinted in* 1964 U.S. Code Cong. & Ad. News 2391, 2401. Consistent with that goal, Title VII prohibits employers from discriminating in employment decisions based on such impermissible classifications. *See* 42 U.S.C. § 2000e-2. However, because "Congress did not intend by Title VII . . . to guarantee a job to every person regardless of his qualifications," *Griggs v. Duke Power Co.*, 401 U.S. 424, 430 (1971), it authorized employers to distinguish among individuals for selection or promotion purposes based "upon the results of any professionally developed ability test provided that such tests, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex or national origin." 42 U.S.C. 2000e-2(h).[44]

In permitting the appropriate use of "professionally developed ability tests" as a basis for selecting and promoting individuals, the drafters of the ultimate language of Title VII stressed the importance of demonstrating the relationship and relevance of the selection procedure to job qualifications. *See* 110 *Cong. Rec.* 7247 (1964). Recognizing the "u[tility]" of "testing or measuring procedures," this Court has also stressed Congress' intent to "forbid[] . . . giving these devices and mechanisms

---

[44] "[T]he Act does not command that any person be hired simply because he was formerly the subject of discrimination, or because he is a member of a minority group . . . . What is required by Congress is the removal of artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of racial or other impermissible factors." *Griggs*, 401 U.S. at 431.

controlling force unless they are demonstrably a reasonable measure of job performance." *Griggs*, 401 U.S. at 436. "What Congress . . . commanded is that any tests used must measure the person for the job and not the person in the abstract." *Id.*

Proponents of Title VII opposed authorizing the use of "professionally developed ability tests" without regard to their ability to predict performance of the particular job in question. 110 *Cong. Rec.* at 13504. The use of tests which, although "professionally developed" bear no relation to the job for which they are being used to assess individuals, was clearly recognized as a potential means of covert "[d]iscrimination[,] . . . under the guise of compliance with the statute." *Id.* Assessment devices which have not been shown to be job-related, or otherwise predictive of performance of a particular job, cannot justify discriminatory employment practices. The crucial public policy goals of Title VII would be thwarted if employers could rebut claims of discrimination simply by pointing to the results of unvalidated assessment devices, whether subjective or objective. Indeed, such unvalidated results may well reflect precisely the discrimination Congress sought to eliminate in Title VII.

In light of the "nondiscrimination" objectives of Title VII and the demonstrated ability of professionals to validate subjective assessment devices, however, there is no principled reason to treat objective and subjective devices differently in imposing a validation requirement, regardless of whether a plaintiff proceeds under disparate impact or disparate treatment claim.[45] Indeed, permitting the use of unvalidated subjective assessment devices while requiring objective devices to be validated provides a ready mechanism for covert discrimination for employers seeking to avoid the constraints of Title VII. Validation does require the expenditure of both time and money by an employer. But, as *amicus* has demonstrated,

---

[45] *Texas Dep't of Comm. Affairs v. Burdine*, 450 U.S. 248 (1981); *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973).

there are a number of readily available techniques for developing, adopting, and validating both objective and subjective devices, and both professional and legal standards allow the use of already developed selection devices. Thus, when one balances the relative costs of validation to the employer against the costs of eroding the protections provided by Title VII and the damage to society of perpetuating the vestiges of discrimination, the outcome clearly favors the requirement that employers use pyschometrically sound and job-relevant selection devices.

## CONCLUSION

For the foregoing reasons, *amicus* respectfully requests that this Court reverse the decision of the Court of Appeals for the Fifth Circuit insofar as it releases employers from their obligation to show that the selection devices they use to make employment decisions are valid.

Respectfully submitted,

DONALD N. BERSOFF
(Counsel of Record)
LAUREL PYKE MALSON
DONALD B. VERRILLI, JR.
ENNIS FRIEDMAN & BERSOFF
1200 - 17th Street, N.W., Suite 400
Washington, D.C. 20036
(202) 775-8100
*Attorneys for* Amicus Curiae
*American Psychological Association*

September 14, 1987